PROSPECTIVE SECONDARY MATHEMATICS TEACHERS' CONSTRUCTION OF BOX PLOTS AND DISTRIBUTIONAL REASONING WITH THREE CONSTRUCTION TOOLS

Samet Okumuş, Emily Thrasher

North Carolina State University

In this study we examined two prospective secondary mathematics teachers' constructions of box plots and their understanding of the distribution they were representing. The participants constructed box plots with paper-and-pencil, graphic calculator and TinkerPlots during clinical interviews. The study indicated that prospective mathematics teachers recognized that using technology to construct box plots provided affordances compared to creating a box plot by hand.

INTRODUCTION

In statistics, data can be represented in many different ways such as graphs and tables that have the potential to provide new understandings of the characteristics of the data (Myatt, 2007). Bakker, Biehler, and Konold (2004) emphasize that box plots provide rich representations since they give information about both measures of center and spread of the data, and can facilitate making comparisons of distributions. Although box plots are viewed as effective representations, it has been documented that students struggle with understanding the data they convey. Box plots can be challenging for students to understand because data is presented as aggregate instead showing individual points and understanding the median and quartiles is not as intuitive as once suspected (Bakker, Biehler, & Konold, 2004). Additionally, delMas (2004) stresses that "understanding how the abstract representation of a "box" can stand for an abstract aspect of a data set (a specific, localized portion of its variability) is no small task" (p. 87).

These problems could be minimized with the availability of technology in statistics. Chance et al (2007) outline many effective uses of technology in the learning of statistics. Three of these categories are automation of calculations, emphasis on data exploration, and visualization of abstract concepts. Automation of calculations allows for timely calculations with high accuracy and emphasis on exploration suggests that many graphs can be produced quickly. Visualization of abstract concepts is the idea that technology helps students to "see" statistical concepts. These uses of technology can potentially help students with the challenges of box plots.

Although there are many statistical packages/technologies that can help students create box plots, two widely used options are *TinkerPlots* (*TP*) (Konold, & Miller, 2005) and graphing calculators (GCs). Burrill (1997) studied the roles and potential of using GCs and remarked that, using GCs, students could be able to see if a data set contained an outlier, which could allow them to exclude the outlier from the data set and reexamine

2014. In Liljedahl, P., Oesterle, S., Nicol, C., & Allan, D. (Eds.) Proceedings of the Joint Meeting of PME 38 and PME-NA 36,Vol. 4, pp. 369-376. Vancouver, Canada: PME.

the distribution. On the other hand, Garfield and Ben-Zvi (2008) found that *TP* allowed students to perceive individual data values of a box plot which facilitates students' understanding. These studies focused on how the technology helped students understand box plots, but it is also important to focus on if teacher notice and appreciate these allowances when working with these technologies.

In this study, we examined two prospective mathematics teachers' thinking about box plot constructions by paper-and-pencil, GC, and *TP*. Using the above-mentioned categories of effective uses of technology (Chance et al., 2007) as a framework, we examined how prospective mathematics teachers reasoned while representing a data set. Accordingly, we identified the challenges and understandings of each prospective mathematics teacher as well as highlighting the teacher's recognition of the affordances of each type of technology.

METHOD

Participants

The participants of this study consisted of two prospective secondary mathematics teachers (1 male, 1 female) who were enrolled in a course about teaching mathematics with technology. These participants were selected based on recommendations from the instructor and their availability to meet with the researchers. Pseudonyms (Amy and John) are assigned to the participants. Both participants were seniors and their ages were 21. Neither participant had experienced using *TP* before the interviews but both had used GC.

Task and Interviews

The task used was taken from the Number of Rope Jumps data (Lappan et. al 2003, p. 40), which describes the maximum number of rope jumps for each student of a 28-person class. The data had a large variation and contained an outlier. Semi-structured clinical interviews were conducted individually with the participants. A *TI-84 Plus Silver Edition* GC, a laptop with *TP* software, a ruler, and paper were provided for the interviews. The data set was already entered as lists in the GC and available as a set of data cards in *TP*.

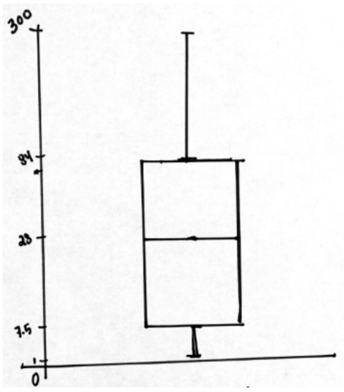
The data for this paper comes from a larger interview about multiple data sets. Each interviewee was asked to construct a box plot using paper-and-pencil first, then a GC, and lastly using *TP*. In addition, the interviewees were asked to construct a box plot after the outlier (300) was excluded from the data set by hand. The interviews, which took about an hour and a half with each participant, were videotaped and voice recorded. The interviews were transcribed and the transcribed data was analyzed descriptively. We analyzed the data by three main categories; which were box plot constructions with paper-and-pencil, using the GC, and using *TP*. In each category instances of reasoning with box plots and the issues or affordances of the technology were identified. Data matrices were constructed (Benard & Ryan, 2010) for each category in order to compare and contrast the interviewee's responses.

RESULTS

Box Plot Constructions with Paper-and-Pencil

At the beginning of the interviews, both interviewees were given the data in a table, and asked what is needed to construct a box plot. Both interviewees mentioned the requirement of a five number summary, and each used 1-Variable stats from graphing calculator to find the five number summary then constructed the box plot by hand. Amy constructed a vertical box plot while John constructed a horizontal box plot (Figure 1 and Figure 2).

In both cases the interviewees provided a number line with a scale but acknowledged that their scales were only estimates and not exact. This is important to note because having imprecise representations makes reasoning about the data more difficult. In fact Amy was aware of her inaccurate scale by saying: “the scale is gonna be kinda off” while constructing her box plot.



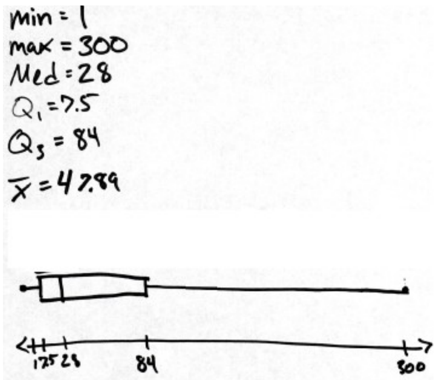


Figure 1: Amy’s box plot Figure 2: John’s box plot

After constructing the box plots, each interviewee was asked to reason about the distribution. Amy (A) said it, “looks like the data is probably skewed. I guess to the right”. Then she explained:

A: ... And if I look at this sheet [data table], I can kinda see. That most of it, 300, is kinda out by itself, but I have like a few 90’s 93, 96 that’s still pretty close to 84 [upper quartile] So, most of the data is right around the median besides this one 300 which is way out here [points to the upper whiskers] Oh, there is a 113 [in data table], that’s OK, that is kinda in there [shows a point on the upper whisker]. But I’d say it is pretty accurate [the box plot] based [on], like, the skewness of it, everything is pretty accurate, but kinda skewed.

Although Amy was able to determine the shape of her distribution, Amy referenced the individual cases from the table to decide on the shape of the distribution instead of using her box plot. This suggested that Amy did not understand how her box plot described the shape of the data or that she is was sure about the accuracy of her representation. When John was asked to reason about the distribution of the data with his box plot (Figure 2), he attended to an aspect of the variability by noting “It is

definitely clustered before the median, before 28 because there is much smaller range between the minimum and the median in that case." Since John did not refer back to the data and used the aspects of a box plot to discuss the distribution, John demonstrated a better understanding of a box plot. However, unlike Amy, John did not address the shape of the distribution.

Next, both interviewees were asked to identify any outliers and create a box plot without the outlier (Figure 3 and Figure 4). Amy had difficulty identifying whether 300 was an outlier. She could not clearly express how she could identify outlier(s) in a data set, and she said that she forgot the formula for identifying an outlier. On the other hand, John did not have such difficulty. He applied the typical 1.5*IQR + Q3 formula of identifying outlier(s) in a data set.

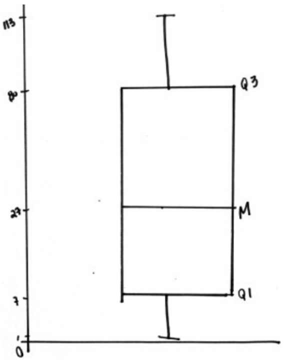


Figure 3: Amy's box plot after excluding the outlier

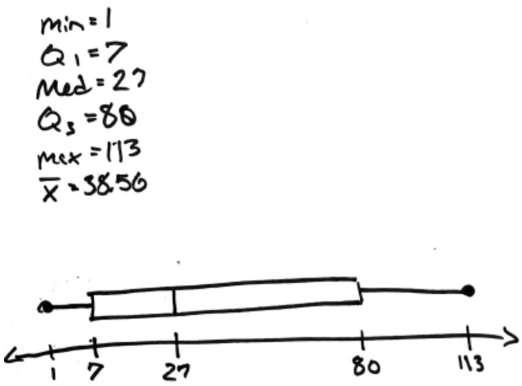


Figure 4: John's box plot after excluding the outlier

While both interviewees were capable of creating a box plot for the data with little trouble, each had their own challenges. First, the constructions with paper-and-pencil were not accurate since both interviewees constructed box plots with a poor scale. Amy demonstrated difficultly with the aggregate nature of the box plot and needed to refer to the individual cases to describe the shape of the distribution. Additionally, Amy had a problem identifying the outlier of the data set. On the other hand, John used the aspects of the box plot to reason about the variability demonstrating a better understanding of the representation.

Box Plot Constructions Using the GC

Next, interviewees were asked to construct box plots using the GC. Both interviewees chose to create a modified box plot with 300 denoted as the outlier. The researcher asked the interviewees to compare these to their previous box plots. Since Amy constructed all vertical box plots with the paper-and-pencil environment and the GC only constructed horizontal box plots, she rotated her paper to view her previous graphs horizontally while comparing them with GC's. When asked how the box plots were alike and different, she answered as follows:

A: Theirs [GC] is much more accurate scale-wise…You can see…how they have it set up scale-wise. You can barely see that little whisker but it is

really close 1 and 7, which is easy to see. And then, you can see that—I mean mine is just more spread out. The scale is much better [GC].

On the other hand, John stressed the differences of his previous and current box plot constructions as follows:

J: It is different because it doesn't, mine doesn't show that there is, doesn't consider the outlier not a part of the actual box plot. So, if I wanted to remedy that, then I would have my maximum here but I would have a lone dot….

Also, the researcher asked the interviewees whether there was anything they wanted to change about their early understanding of the data after they constructed a box plot using the GC. John stated that "no my understanding stayed pretty, pretty [un]impacted. But it is nice to immediately be able to tell about outliers instead of having to calculate them for myself". On the other hand, Amy believed that her understanding changed a lot. She addressed the accuracy of construction of the GC saying "this'd [showing the GC] tell a lot more versus this [showing her box plot]. This is, just looking at it [hand drawn box plot] looks more deceiving whereas this one [GC] it's very accurate…"

In both cases, the calculator's ability to quickly and accurately create a box plot was considered helpful. For Amy, the accuracy of the scale helped her to better understand the distribution. In fact, she believed that her hand drawing was misleading. John appreciated the ability of the calculator to find and denote the outlier quickly.

Box Plot Constructions Using *TP*

For the final box plot, the interviewer constructed a box plot (Figure 5) within *TP* because interviewees were not familiar with the tool. Interviewees were asked to reason about the data and compare this graph with the previous box plots. The first impression of John about the representation was as follows:

J: This does not consider outliers although I would assume that we could make it consider outliers. This is really, really nice being able to show or because it shows where the data points lie. So you can clearly see that the data is clustered average of the left side (rope jumps) and as you and as you go to the right there is less and less data except, except for like around 90 apparently they get tired at 90. Yeah this is really useful.

Interestingly this is also the only instance in both interviews where the jump rope context was addressed.

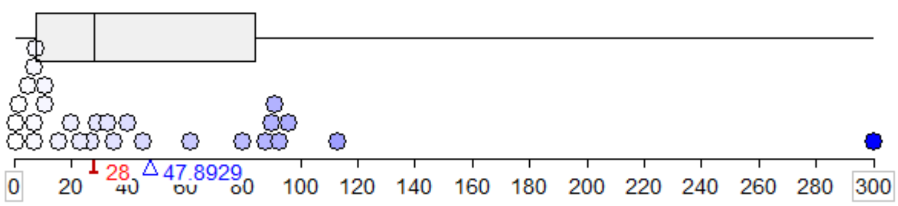


Figure 5: The box plot representation in TP

Compared to the GC representation of the box plot, John thought *TP* was superior since it could show the box plot as well as the individual data points. This gave him information to better understand his data and changed his understanding:

J: I think that this actually did change [his understanding] a little bit. We can see there is another cluster around 90, but with the other methods that I have been using there's no way to turn that just by looking at it, you have to actually examine the data itself.

In recognizing the there was another cluster around 90 that the box plot did not show, John was also gaining a better understanding of the pros and cons of the box plot representation. Amy provided similar expressions in terms of visualization of *TP* and being able to see individual data points, and addressed that the box plot representation in *TP* is better than the previous representations.

A: This, this is awesome [TP]…I know there was a lot of people that I was in school with that struggled and probably this kinda software--probably help them, a lot, with learning box plots and so, just visually you can see it. I mean I like this (showing the GC) and obviously, like, paper-and-pencil, probably, you can do it but it is not as effective. I mean like, like as I said earlier, like, mine were kinda skewed (refers her all previous box plots using paper-and-pencil) and you could see, when you saw it on the calculator and then, like, putting this information to this (showing *TP*) it's just even more--visual. I really liked it.

At the end of the interview, the researcher asked for the interviewee's last comments about the three different technology representations. John's thoughts summarize many of the affordances of each technology.

J: I'll start with the paper. The pros, you can very clearly tell whether or not a student understands what minimum, umm the quartiles, the median, and the max represent…The paper method does make it more difficult to recognize when there are outliers, however. Calculator, it is nice in most cases because you're able to use the trace to tell where each important area is--quartiles, and median, minimum, max. But, there is not really, there is not necessarily the understanding of what is going on behind the graph. You can't tell for sure whether the students know without talking to them directly and in a large class that (inaudible) confusing and hectic. As for *TP*, it's pretty much, I can't come up with any cons, but it's, it's really nice to be able to see the data points to see where they lined at any point in time, it gives a really good visual representation of what each of the four areas or four quadrants represent.

DISCUSSION AND IMPLICATIONS

This study indicated how prospective mathematics teachers reasoned about distributions with box plots while using different tools. Amy initially had trouble understanding the distribution of the data set without the use of the individual cases while John demonstrated a better understanding of the box plot representation. In both

cases, the use of technology changed their understanding of the distribution. For Amy, technology provided an accurate representation because her scale was poor. Additionally, Amy and John found that having the individual cases in *TP* gave them new insights into their distribution. Finally, John explicitly expressed that having the outlier identified and marked on the box plot was helpful to him. Accordingly, we could conclude that these prospective teachers recognized the ability of technology to create box plots accurately and with additional information (denoted outliers or individual data points).

Interestingly, all three of the aforementioned Chance et al (2007) categories for effective uses of technology were observed by the prospective teachers. Both interviewees recognized the strengths of using technology for creating graphs accurately and quickly, automation of calculation and emphasis on data exploration were acknowledged by the prospective teachers. Finally, since box plots are an abstract representation (delMas, 2004) and the teachers expressed appreciation for technologies ability to help visualize the box plot, the prospective teachers are recognizing technologies ability to visualize abstract concepts.

Although more research needs to be conducted in this area because of the small sample size of this study, the study findings suggest that prospective teachers should have experience with different types of technology to produce box plots. This exposure may help to produce prospective teachers that both develop deeper understanding of box plots and that are more likely to use different type of technology in their future classrooms.

References

Bakker, A., Biehler, R., & Konold, C. (2004). Should young students learn about box plots? In G. Burrill & M. Camden (Eds.), *Proceedings of IASE 2004 roundtable on curricular development in statistics education* (pp. 163-173), Lund, Sweden. Voorburg, The Netherlands: International Statistical Institute.

Bernard, H. R., & Ryan, R. W. (2010). *Analyzing qualitative data: Systematic approaches*. Thousand Oaks, CA: Sage Publications.

Burrill, G. (1997). Graphing calculators and their potential for teaching and learning statistics. In J. B. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics: Proceedings of the 1996 IASE round table conference* (pp. 15-28). Voorburg, The Netherlands: International Statistical Institute.

Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education*, *1*(1), 1-26.

del Mas, R. (2004). A comparison of mathematical and statistical reasoning. In D. Ben- Zvi & J. B. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 79-95). Dordrecht, The Netherlands: Kluwer.

Garfield, J. B., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: connecting research and teaching practice.* The Netherlands: Springer.

Konold, C., & Miller, C. D. (2005). *TinkerPlots: Dynamic Data Exploration*™ [computer software]. Emeryville, CA: Key Curriculum Press. Retrieved from http://www.keypress.com/x5715.xml.

Lappan, G., Fey, J. T., Fitzgerald, W. M., Friel, S. N., & Philips, E. D. (2003). *Connected mathematics data about us statistics* (student edition). Lebanon, IN: Dale Seymour Publications.

Myatt, G. (2007). *Making sense of data: A practical guide to exploratory data analysis and data mining*. Hoboken, NJ: Wiley-Interscience, John Wiley and Sons, Inc.